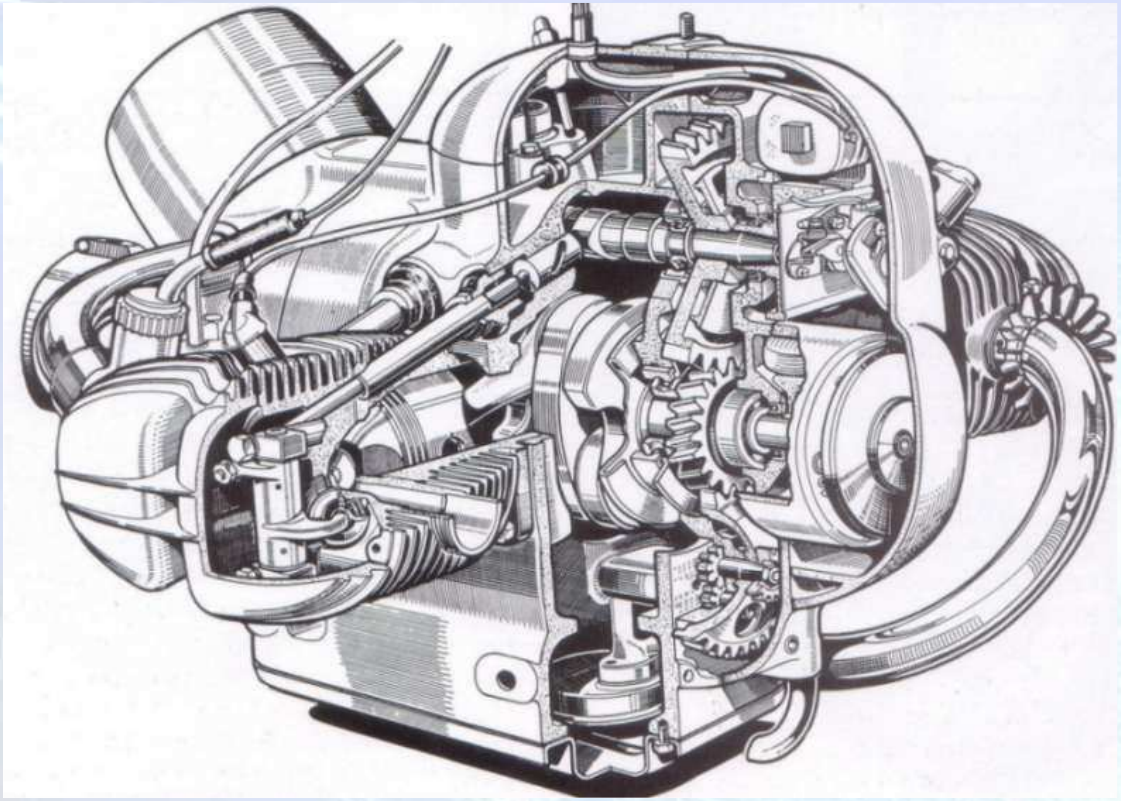
Statistical Engine and Central Engine in pratice



(The Working Example)

The implementation of the Statistical Objects

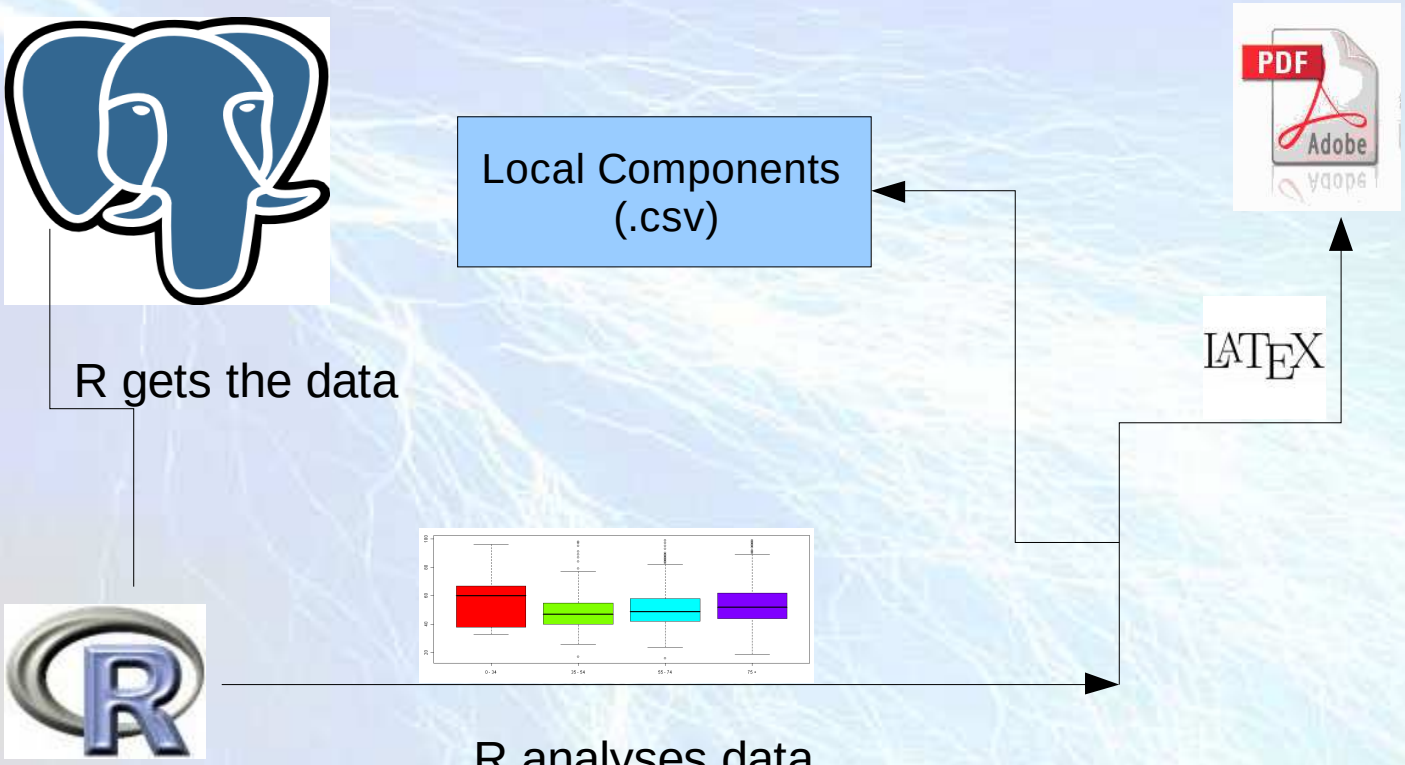
Luca Rossi

Objective of these slides

- The objective of these slides is to illustrate a working example of the statistical engine and the central engine, paying attention to the implementation of the statistical objects.

The most part of this presentation is done executing the software (the working example)

Briefly the process...



This structure is adopted both locally and at the central level

What's the Statistical Engine?

- The S.E. is a set of functions written in R, that combined allows us to produce analysis, report and csv file in predefined formats
- R is not an high level programming language, so we have written a lot of functions to satisfy our requests.
- We can divide functions in 3 groups:

1. Functions for the report (texfile and html)
2. Functions based on the statistical objects (analysis)
3. Functions for the selection of cohort

The step-by-step example

1.Connection through JDBC driver (with the R package RJDBC)

```
library(RJDBC)
drv <- JDBC("org.postgresql.Driver","F:\\biroR\\jdbc\\postgresql-8.2-504.jdbc3.jar", "`")
conn <- dbConnect(drv, "jdbc:postgresql://localhost/centro5",user="postgres1",password="luca",dbname="centro5")
```

2.Importing data from Pgsql db, through this script

```
## Import data from Postgres

patvar<-c("SEX","DOB","DT_DIAG","TYPE_DM")

patient<-as.data.frame(dbGetQuery(conn, "Select DISTINCT patient_id FROM profile"))

Q<-rep(NA,length(patvar))
for (i in 1:length(Q)) {
 Q[i]<-paste("Select DISTINCT profile_field_value AS ",patvar[i],",patient_id  FROM profile WHERE profile_field_name='",patvar[i],"'",sep="")
}

for (i in 1:length(Q)) {
 temp<-as.data.frame(dbGetQuery(conn, Q[i]))
 patient<-merge(patient,temp,by="patient_id",all.x=T)
 rm(temp)
}

rm(Q)
centre_id=as.character(unique(dbGetQuery(conn, "Select site_header_id from site_header"))[1])
```

The Step-by-step example

Clinical data are loaded (in RAM) in two datasets (data.frame in R) and then stored in a specific directory (csv format):

. one called “patient” in which we have “patient_id”,”sex”,”dob”,”type_dm”,”dt_diag”

. and one called “episode” in which we have “patient_id”, “episode_date”, “weight”, “bmi”, “hba1c”, “ma_test”,...
. then we have other data.frame for mortality, and population data, variable for centre_id,...

After that we have:

Conversion of some variables

Creation of categorical variables, according to predefined thresholds (those, in the alpha version of the S.E., will be in a configuration file)

Let's go to see the first indicator

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
            caption="Barplot: Age by Gender",
            width=1,
            dirgraph=dirgraphlatex,
            texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

Let's go to see the first indicator

Testing conditions of the Feasibility

We have the distrib[illegible] age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
           caption="Barplot: Age by Gender",
           width=1,
           dirgraph=dirgraphlatex,
           texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

Let's go to see the first indicator

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

 data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
 #data1.2 is the same of data1.1

 BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
 lev=list(a=classlabellist(ageth),b=levsex),
 texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
 caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

 BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
       lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
       number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

 import_png(namegraph="bar1_1.png",
          caption="Barplot: Age by Gender",
          width=1,
          dirgraph=dirgraphlatex,
          texfile=texfile)

 new_sub_section(title="Gender",texfile=texfile)

 BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
 lev=list(a=levsex,b=classlabellist(ageth)),
 texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
 caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

 rm(data1.1) #Remove del data.frame from the workspace
}
```

BiroDemographic is a function that allows us to calculate age, at a particular pre-defined date like
2006/12/31,
at the last episode date,
at the analysis date,
at the last episode date of a patient for each year,
at the first episode date
at each episode date

in this case at the analysis date

this function return a data.frame with the variables specified in the “var” parameter, eventually with also “patient_id” and <date> if the “hold” parameter id equal to TRUE

Let's go to see

We have the distribution of the age of the p

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),dat
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
           caption="Barplot: Age by Gender",
           width=1,
           dirgraph=dirgraphlatex,
           texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

This is a function that
manages the layout of the table, tha calls another function that create a contincengy table, write a csv file with the values,
this last calls another function that print the table in a Tex file and an html table (html: not completed)

Let's go to see the first indicator

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
       lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
       number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_pn
          capt
          widt
          dirg
          texfi

  new_sub_s

  BIROtable
  lev=list(a=
  texfile=tex
  caption="C

  rm(data1.1
}
```

Calculate the height of the bars (and write csv), and calls the fanction that realizes graphs (jpg,epg,png format)

Write in the tex file graphs that must be included

Let's go to see the first material

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
           caption="Barplot: Age by Gender",
           width=1,
           dirgraph=dirgraphlatex,
           texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

Let's go to see the first indicator

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
            caption="Barplot: Age by Gender",
            width=1,
            dirgraph=dirgraphlatex,
            texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),FirstLab="Gender",
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

Create a sub section in the tex file

Let's go to see the first indicator

We have the distribution of the age of the patients, by their gender

```
if (length(na.omit(patient$sex))>0 && length(na.omit(patient$dob))>0) {

  data1.1<-BiroDemographic(patient=patient,episode=episode,var=c("sex","age_c"),date="analysis date",startvar="dob",th=ageth,hold=FALSE,varname="age")
  #data1.2 is the same of data1.1

  BIROtable(var=list(a=data1.1$age_c,b=data1.1$sex),FirstLab="Age in classes",
  lev=list(a=classlabellist(ageth),b=levsex),
  texfile=texfile,codtab="1_1",dirhtml=dirhtml,dircsv=dircsv,
  caption="Age (by Gender)",side=0,width=0.5,perc=TRUE)

  BIRObar(var=list(a=data1.1$age_c,b=data1.1$sex),namevar=c("Age",""),
        lev=list(a=classlabellist(ageth),b=levsex),beside=TRUE,
        number=TRUE,perc=TRUE,namegraph="1_1",dirgraph=dirgraph)

  import_png(namegraph="bar1_1.png",
           caption="Barplot: Age by Gender",
           width=1,
           dirgraph=dirgraphlatex,
           texfile=texfile)

  new_sub_section(title="Gender",texfile=texfile)

  BIROtable(var=list(a=data1.1$sex,b=data1.1$age_c),F
  lev=list(a=levsex,b=classlabellist(ageth)),
  texfile=texfile,codtab="1_2",dirhtml=dirhtml,dircsv=dircsv,
  caption="Gender (by Age)",side=0,perc=TRUE,width=0.5)

  rm(data1.1) #Remove del data.frame from the workspace
}
```

rm is an R function that removes an R object from the workspace (RAM)

Partial Boxplot

- In the last meeting you have seen the function BIRObox, that created statistics and plotted boxplots.

- Now, implementing the statistical object “Partial Boxplot” we had to split this function in 2 parts:

    1. one to calculate statistics and to save the csv file
    2. one to rebuild fron the csv file an R list to plot the boxplot.

Overall Boxplot

The overall boxplot require 2 functions:

- BIROfd (The implemetation of the frequency distribution, overall boxplot at the local level)
- Overbox (The implementation of the overall boxplot at the central level)

F:\birowe\software_se_\source\R\include\birofd.R

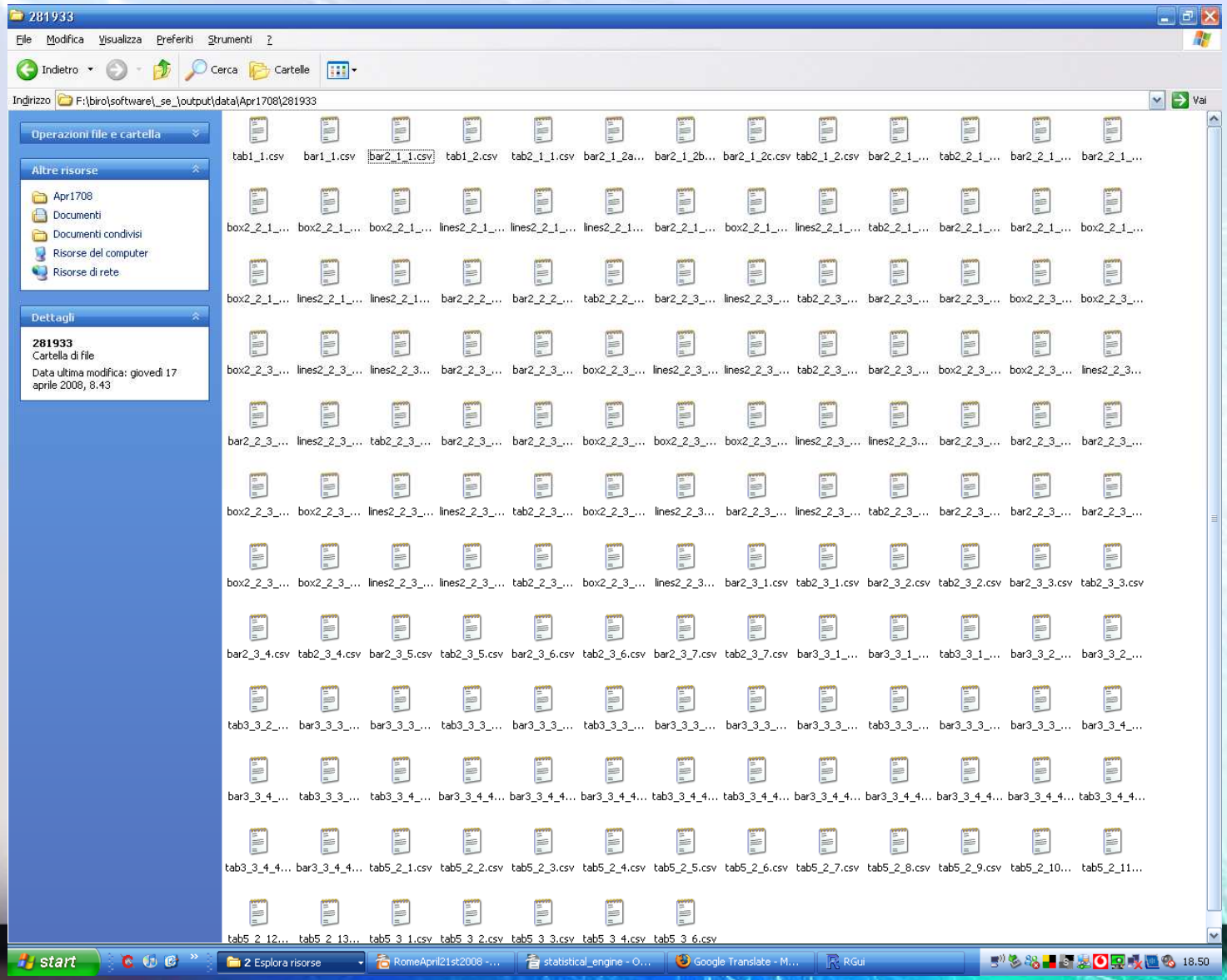
F:\birowe\software_se_\source\R\include\overbox.r

Others

We have implemented also other statistical objects like:

- arithmetic mean
- var
- range
- ...

Now all the csv are in a specific directory



Now these csv are sent to the server

- With the comunication software
- Stored in a specific directory

(These there is a Java program that controls if there are new data and activates the CSV Importer)

- Loaded with the CSV Imported (A variant of the BIRO Adaptor) into a Postgreqsl database

Now we simulate all these steps using a batch command that calls the CSVImporter directly from the (“local” output directories)